

Covid-19 MLIA task 3

Francisco Casacuberta¹, Miguel Domingo¹, Mercedes García-Martínez²,
Manuel Herranz²

{fcn, midobal}@prhlt.upv.es,
{m.garcia, m.herranz}@pangeanic.com

¹PRHLT Research Center - Universitat Politècnica de València

²Pangeanic / B.I Europa - PangeaMT Technologies Division

CHIST-ERA Challenge Call

Online, July 21, 2021

Outline

1. MT task
2. Submissions
3. Results

Outline

1. MT task
2. Submissions
3. Results

MT task

The goal of the MT task is to generate MT systems focused on Covid-19 related documents for different language pairs. It is composed of three rounds.

Examples:

- 30% of children and adults infected with measles can develop complications.
- The MMR vaccine is safe and effective and has very few side effects.

Language pairs

- English–German.
- English–French.
- English–Spanish.
- English–Italian.
- English–Modern Greek.
- English–Swedish.
- **English–Arabic.** New for round 2.

Categories

- **Constrained:** systems which have been trained exclusively with data provided by the organizers (compulsory).
- **Unconstrained:** systems which have been trained using external data and/or resources (optional).

Corpora

- Data from well-known web sources from health and medicine domains.
- COVID-19-related data from other web sources (e.g., broadcast websites).
- Generated by the data acquisition team.

Evaluation

- BLEU*.
- ChrF.
- **TER** and **BEER**. New for round 2.

*Main metric.

- Approximate Randomization Testing (ART)^{1,2}.

¹Riezler, S., Maxwell, J.T.: On some pitfalls in automatic evaluation and significance testing for mt. In: Proceedings of the workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization, pp. 5764 (2005).

²github.com/midobal/covid19mlia-mt-task/tree/master/round1/art.

Quality assessment

- 500 Spanish segments post-edited by a team of professional translators.
- 18.8 TER from the reference and its post-edited version.
- Evaluation is consistent using the reference and its post-edited version.

Outline

1. MT task
2. Submissions
3. Results

Submissions

- Baselines:
 1. Standard Transformer.
 2. Standard RNN.
- 8 different teams took part in this round.

Outline

1. MT task
2. Submissions
3. Results

Results

- The most successful approaches were based on multilingual MT and transfer learning.
- PROMT's approach:
 - ▶ Best results for all language pairs in both categories except for unconstrained English–German and constrained English–German sharing the first position with CUNI-MT.
 - ▶ Multilingual system trained using all language pairs, a smaller vocabulary and sentencepiece.
- In general, the difference between systems from one rank and the next one is small. The RNN baseline delimits the point in which there is a significant drop of translation quality between ranks.

Challenges

- We detected an anomaly in the results of the unconstrained category: systems trained exclusively with WMT news data performed better than systems trained with medical data.
- Solution for round 2:
 - ▶ Make sure to balance the test and validation sets properly.
 - ▶ Study the data in more detail to remove anomalies.